

# Visual Protocol Collection for the Enhancement of the Radiological Diagnostic Process

E. Rogers<sup>1</sup>, R.C. Arkin<sup>2</sup>, M. Baron<sup>3</sup>, N. Ezquerro<sup>4</sup>, E. Garcia<sup>5</sup>

## Abstract

This paper describes the Observation Phase of a research project that is designed to explore the nature of visual radiological diagnosis. It is concerned with the collection and analysis of visual protocol data in an effort to determine the salient image features that may impact on the different stages of the diagnostic process, both perceptual and cognitive. These features, once identified, will lead to the formulation of image enhancement hypotheses and the development of knowledge-based visualization techniques to assist the radiological diagnostic process.

## 1 Introduction

The application of artificial intelligence techniques to problems in the area of biomedical visualization is a relatively young field of endeavor. An important overall goal of this research is the identification of fundamental principles and concepts with which to define and describe the mechanisms of visualization. Of particular interest is the visualization of two-dimensional medical image data since the specialists who work in this domain are trained to extract information that is *already in a visual format*. Investigation of such specific "visual interactions" should facilitate the study of the entire visualization process.

A joint research project currently underway at the Georgia Institute of Technology and the Emory University School of Medicine addresses some of these issues. It is designed to explore the nature of visual radiological diagnosis and has two main objectives:

1. To develop an understanding of how radiologists interpret medical images in the process of solving diagnostic problems.
2. To devise and implement knowledge-based visualization techniques in order to improve diagnostic accuracy and throughput.

The relationship between the raw data (in this case, images) and the radiologist's use of these data is explored to determine: how much of the extracted information is inherent in the raw data itself; what is the nature of this extracted and perceptually processed information; how much information is produced by the cognitive processes of the radiologist acting upon the raw data; and if it is possible to utilize suitable image transformations to facilitate the interpretative and diagnostic processes in light of our understanding of the active perceptual and cognitive processes used.

To meet the first objective of the project, three data collection phases have been identified: an *Observation Phase*, an *Intrusive Phase*, and an *Enhancement Phase*. The Observation Phase, designed to non-intrusively collect information on how radiologists normally interact with images, provides baseline data for the subsequent phases, and is subdivided into two parts: *Reading Room Observation* and *Transitional Observation*. Reading Room Observation involves the study of the regular procedures of a chest X-ray reading room in a hospital, in order to observe radiologists in their normal environment. The Transitional Observation part has three purposes: first, to acclimatize the radiology residents and experts participating in the study to the environment

of the laboratory where the subsequent phases of the project will be conducted; second, to allow the subjects to make the transition from reading chest X-ray images on film, to reading chest X-ray images on a computer screen (CRT); and third, to compare the visual protocols resulting from both film and CRT interaction with respect to features, timing and performance. This information is essential to the design of the Intrusive and Enhancement Phases, where only CRT images will be used.

The Intrusive Phase is so-called because the investigator takes an active role in eliciting information from the subjects while they are looking at CRT images of chest X-rays and it is expected that the results from this phase will lead to the formulation of three kinds of hypotheses:

1. Classification and description of features requisite for diagnosis that are inherent in the images.
2. Preliminary modeling of the cognitive processing that occurs during visual diagnostic activity.
3. Formulation of image enhancements designed to assist and improve performance in the diagnostic task, especially on the part of non-experts.

These hypotheses will then be tested in the Enhancement Phase, which will examine the subjects' diagnostic performance with various kinds of enhanced CRT images. Four key issues addressed in this phase are:

**Dimensionality:** determination of the nature and quantity of the information that is present in the image and required for diagnosis.

**Expectation:** the impact of case history information and prior experience on how the subject sees the image.

**Context:** the role of particular spatial configurations of image events and the implications of perceptual organization of landmarks and visual cues.

**Focus-of-Attention:** the channeling of perceptual and cognitive resources to the regions of the image central to the diagnosis.

This paper is concerned with presenting the results of the Observation Phase of the project with a view to designing the next set of experiments for the Intrusive Phase. The analysis of this visual protocol data aims to determine salient image features that merit more detailed examination. An important aspect of this analysis is the exploration of distinctions between resident and expert performance, which may provide signposts for particular types of enhancements. Although the initial domain for our work is chest X-rays, it is expected that the visualization principles which evolve will be extendable to other areas of radiology as well.

The remainder of this paper is subdivided into the following sections. Section 2 surveys related work. Section 3 describes the methodologies used for data collection. Section 4 presents an analysis of the data. Section 5 discusses conclusions and future work, and a summary completes this report.

## 2 Related Work

An overview of current literature reveals that considerable effort has gone into the study and development of computerized image analysis, image enhancement and automatic detection. An examination of these results reveals certain important features and problems encountered in attempts to optimize physical properties of images with a view to improving diagnostic performance.

There have been several attempts to link the ability of readers to detect nodules on chest X-rays with the physical properties of the image. Revesz and Kundel [17] have examined nodule properties such as size, shape and edge gradient coupled with background properties including overall density and complexity. The aim of Sezan et al [20] was to optimize the appearance of the lung fields and the mediastinum, two regions which have different enhancement requirements, by selectively performing gray-scale modification and unsharp masking. Van der Wildt and Waarts [26] addressed the influence of surround on contrast sensitivity and other researchers such as Kimme and Sklansky [9], Ligtenberg et al [12], and 't Hoen [24] have also studied noise, contrast and unsharpness effects on image quality. Sharp et al [21] noted the effects of pixel size on performance related to radionuclide images, and also studied observer preferences with respect to the varying pixel sizes. Webb [27] expanded on the notion of pixel significance and suggested that varying the pixel size to suppress insignificant detail may improve the information content of images.

Systems that perform automatic detection and classification of various abnormalities of the chest such as lung nodules [22], prepulmonary edema [1], interstitial lung disease [8], and general chest abnormalities [28] focus primarily on one type of feature, such as vascular patterns, texture or shape and attempt to produce a computer system that can extract and classify this single feature. Performance of these systems is often tested only against previously diagnosed cases, and does not necessarily lead to a better understanding of how humans extract such features and achieve a diagnosis.

Hendee and Wells, in a recent article [7], discuss three common models for understanding how observers detect information present in visual images: the signal-to-noise model, which is based on the detection of simple visual signals by an observer; the computational model, which proposes that people use a gestalt process to produce a "primal sketch" of an image for rapid initial feature extraction; and finally, the vision channel model, which describes the human visual system as a collection of frequency channels, each tuned to information of a specific type (e.g., configuration, chromaticity, contrast and color). This framework of models is used to guide the next part of the discussion.

A number of researchers deal with observer performance in nodule detection tasks. Moseley and Kelsey [14] tried to measure several aspects of observer performance with respect to nodule detection in radiographic images, including experiential differences, viewing distance effects, stereoscopic pair viewing and direction of attention. In another study of this type, Starr et al [23] described a model for predicting observer performance in an experiment requiring both detection and localization of a lesion, thus adding some complexity to the visual task.

The computational model, developed primarily by Marr and Nishihara [13], seems particularly suited to describing the findings of some visual search studies. Kundel and Nodine [10] tried to determine how much information is extracted in a single glance as compared to prolonged search and found support for the hypothesis that the initial glance provides much basic information. Christensen et al [4] also concluded that the perception process can be split into a rapid and a slow component. In the rapid phase, obvious abnormalities are detected almost instantaneously.

In the third category of vision models is the work of Baker and Lloyd [2], who developed a system of bandpass filters for image processing based on the human vision channel model, and Piotrowski [16], who referred to the spatial-frequency theory of human vision in his work on the importance of spatial-frequency phase information for human recognition and classification.

Models of the more cognitive attributes of vision have also been proposed. In 1962, Tuddenham [25] emphasized the importance of memory and "search for meaning" in the construction of what an ob-

server perceives. This approach led to Blesser and Ozonoff's later work [3] on a three-stage model of perception in diagnostic radiology, including psychophysical, psychological and nosological components. More recently, the work of Lesgold et al [11] provided the foundations for a cognitive model of expert radiological behavior which included the activation of general diagnostic schemas, constraint-posting, opportunistic planning, mental representations of patient anatomy and recursive refinement.

Closely coupled with the above concerns is the exploration of expert/novice differences, which many believe may lead to insights into overall problem-solving behavior. The Lesgold research cited earlier [11], emphasizes the role of experience in the successful construction and manipulation of a rich mental representation of patient anatomy. However, Myles-Worsley et al [15] found that although memory for abnormal X-ray films, together with detection and processing capabilities of distinguishing features, increased with radiological experience, recognition memory for normal films actually *decreased* with experience. Christensen et al, in their visual search study [4], also indicate that experts perform better than novices in speed and detection of abnormalities. However, the study also indicates that experts tend to stop reading at a time when they are still finding important abnormalities in the film.

The work presented in this paper touches on aspects of all of these related areas. The physical attributes of images provide important basic guidelines for enhancement. However, radiological diagnosis involves complex visual tasks with a multiplicity of features, combined with cognitive activities that direct and monitor these tasks. Further, there appears to be a gap between the psychophysical types of research, which are often conducted using computer displays of images, and the more cognitively-oriented research, which focuses primarily on film images. The current study aims to provide a bridge between the perceptual and cognitive aspects of radiological diagnosis, and to facilitate the transition from reading films to reading CRT images.

## 3 Data Collection

The two parts of the Observation Phase, outlined in the Introduction, are described more fully in the following subsections. The Reading Room Observation took place in the chest X-ray reading room of Emory University Hospital, and the Transitional Observation was conducted in the Imaging Sciences Laboratory of Emory University Hospital.

### 3.1 Reading Room Observation

The purpose of this part of the Observation Phase was primarily to acquaint the investigator with the methods and behaviors of radiologists in the normal environment of the X-ray reading room (Fig. 1). Such familiarity was invaluable in attempting to reproduce similar conditions for the transitional part of the data collection. Since Emory University Hospital is a teaching facility as well, the reading room routine allowed observation, not only of radiologists at work, but also of the interaction between radiology residents and senior faculty.

Normally a resident is assigned to the reading room for a month. During this time, he/she is required to handle the regular case load of chest X-rays that is generated by the hospital. The films are displayed on a large rotating light drum, and each case usually includes frontal and lateral views of the chest, together with any previous films the resident considers relevant. The resident is expected to examine these cases and then to call the senior faculty member on duty that day for a joint inspection of the cases. Finally, the resident again reviews each film (alone), this time dictating the findings and diagnoses. Transcripts of these dictations are typed and then signed by both the resident and the senior faculty member. Of course there is some variation of this routine between residents and also between senior faculty members.

Over a period of three months, the investigator spent several days each month observing the reading room procedure. This was done in two ways. The first few days of each monthly session were spent just quietly sitting in a corner of the reading room as unobtrusively

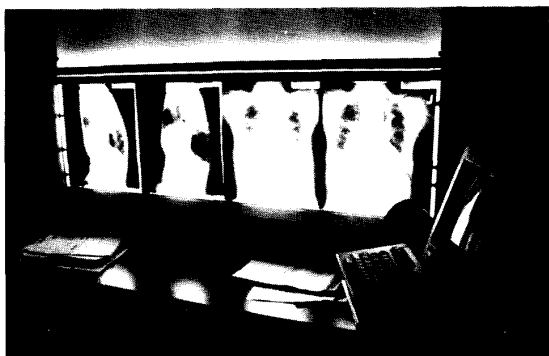


Figure 1: Observation Phase Environment - Reading Room

as possible. Often it was necessary to briefly explain the purpose of the visit to the current resident and/or senior faculty member, but a minimum of details was provided in order to avoid affecting the regular behavior patterns. After this initial accommodation period, the video camera was set up directly behind the radiologists' seats and the morning sessions were taped. Almost without exception, after the first day or two of videotaping, the radiologists appeared to be unaffected by the presence of the camera and the investigator. This data was collected for archival purposes only, primarily to allow the investigator to review and study the regular reading room procedures in preparation for the transitional part of the Observation Phase.

### 3.2 Transitional Observation

This part of the Observation Phase was conducted more formally than the previously described reading room procedure, and consisted of pre-selected subjects reading chest X-ray images in three separate sessions conducted in the Imaging Sciences Laboratory (Fig. 2). The first session was devoted to reading ten X-ray images on film; the second session involved reading ten X-ray images on CRT; and the final session again entailed reading ten X-ray images on film.

#### 3.2.1 Subjects

Ten subjects were recruited from the Radiology Training Program at Emory University Hospital. Two subjects from each year of residency were selected in addition to two staff radiologists of many years' experience. All subjects were qualified M.D.'s and in addition, the second, third and fourth year residents also had at least one year of internship beyond medical school. The two staff radiologists each had approximately thirty years of radiology experience in addition to their medical training and other medical experience. The subjects were first contacted by telephone, and then a brief interview was arranged during which the investigator presented an overview of the study (both oral and written) and the subjects were asked to study and sign a Human Subjects Consent Form to indicate their willingness to participate.

#### 3.2.2 Materials and Design

A total of fifty chest X-ray films were selected for the overall project by our participating radiologist-investigator. The collection includes some standard "textbook-type" abnormal cases, which should be fairly obvious visually to all levels of expertise, some normals, and a number of rather subtle cases that might require some skill to detect the features and to achieve the correct diagnoses. All fifty films were taken to the Division of Imaging Physics at Georgetown University Medical Center in Washington, D.C., and digitized on a Konica X-ray digitizer. The resulting digitized images had a spatial resolution of  $2000 \times 2430$  pixels and a contrast resolution of 10 bits.

For the Transitional Observation part of the study, ten cases were selected randomly from the fifty test cases. Although a variety of com-



Figure 2: Observation Phase Environment - Imaging Sciences Lab

puter equipment was available for displaying the digitized X-rays, none proved optimal as far as spatial and contrast resolution was concerned. Certain studies (e.g., [6], [18], [19]), have indicated that a spatial resolution of  $1024 \times 1024$  is adequate for reading digitized X-rays, although not optimal, and that there may be some degradation of performance if a  $512 \times 512$  image is used. However, due to the varying media and contrast resolutions used in these studies, a more precise lower bound for performance has not been established. For this particular study, it was felt that the long-term goals would be better met by exploiting the maximum available contrast resolution, and thus the Sun/Pixar was chosen as the display medium. Although the Pixar only has a screen resolution of  $768 \times 1028$  pixels, it provides 11-12 bits of gray level contrast. In order to provide some continuity between computer and film images, it was decided to display square-shaped images ( $768 \times 768$  pixels) on the computer screen, and likewise, to reduce the X-ray films to  $14 \times 14$  inches by masking three inches of the film. This was also used to remove any identifying marks from the cases.

To produce the final displayed image on the computer screen, the following processing steps were taken. First, from the original  $2000 \times 2430$  pixel image file,  $2000 \times 2000$  pixels were read into the frame buffer. The extra 430 pixels were just "cut off" strategically from the top and bottom of the image file to remove identifying marks. The 10-bit pixel values were then multiplied by two to conform to Pixar's 11-bit display capability, and the entire  $2000 \times 2000$  image was compressed to  $768 \times 768$  pixels. Finally a reversal procedure was needed, since the digitized images are black-on-white, the opposite of X-ray films. Before performing the reversal, the pixel values were translated so that the maximum corresponded to 2047. This adjustment was necessary because the digitizing process produced an image whose values range between 9 and 10 bits. We thus ensured that, after reversal, the background of the image was black, while maintaining the original histogram shape. No other processing was performed on the images beyond the minimal required to produce a display. We wished to test the subjects with non-optimized images first in order to maximize the potential for improvement in later parts of the study.

In addition to the ten test cases available on film and CRT, ten supplementary film-only cases were included. The test films were mixed with the supplementary films and then split into two groups of ten films each. One set of ten films was viewed by the subjects in the first data collection session. The second session, held a week or more after the first session, consisted of the subjects viewing the ten test cases on the computer screen. Finally, in a third session, again held a week or more after the second session, the subjects viewed the final set of ten film cases. This design ensured that the subjects would have viewed only some of the actual test films *before* seeing them on the computer screen, and some of them *after* the computer session. By coupling this strategy with the time lag between sessions, we hoped to minimize learning and memory effects, so that a reasonably unbiased comparison could be made between performance with film as compared to performance with the computer display.

### 3.2.3 Overall Methodology

The Imaging Sciences Laboratory, site of all three transitional data collection sessions, is equipped with a mounted lightbox (suitable for viewing a single row of radiological films at a time), and several computer display devices, including the Pixar/Sun system. Although the film viewing environment was not optimal compared to the actual radiology reading room, conditions such as low lighting, reduced ambient noise, and minimized distractions were emulated as far as possible. Part of the intention of this part of the Observation Phase included acclimatizing the subjects to working in the laboratory and it was felt that conducting both film reading and CRT reading under similar conditions would be most conducive to that end.

### 3.2.4 Film Viewing Methodology

Both film viewing sessions (i.e., before and after the CRT viewing session) were conducted in the following manner. Each subject entered the laboratory and was seated in front of the mounted viewbox. A demonstration X-ray was placed on the viewbox and the subject was allowed to select a comfortable viewing distance from the X-ray. The video camera was situated several feet behind the subject at an angle overlooking the right shoulder and a microphone attached to a small tape recorder was handed to the subject for dictation purposes.

The subject was told that the session would consist of reading ten chest X-rays, and that, as far as possible, he/she was to behave as though this were a normal reading room procedure where a film is examined and then findings and diagnoses are dictated. There was no time limit placed on the readings, and the subject was to indicate when the reading was finished and the next film was desired. (The investigator replaced the films on the light box, so the subject was not required to move from position.) The subject was advised that the only case information that might be available was age and sex of the patient, and that these would be provided upon request. All camera and video recorder controls were behind the subject and therefore not visible during the reading of the films. Minimal verbal exchange was maintained by the investigator once the session was started and all ten cases were reviewed in succession without interruption.

### 3.2.5 CRT Viewing Methodology

Each subject again entered the laboratory used for the film viewing and this time was seated in front of the Pixar display screen, which was on a table at the opposite end of the laboratory to the lightbox. The investigator was seated at a corner of the same table, with the Sun workstation screen orthogonal to the subject. (This was to minimize distracting light and computer activity.) The ten processed digitized cases were accessed one at a time by the simple pressing of one key by the investigator, again minimizing any extraneous activity. The video camera and recorder were again placed behind the subject, overlooking the right shoulder, and the microphone of the tape recorder was given to the subject for dictation purposes. Conditions were as similar as possible to those maintained during the film viewing sessions including the same instructions, and minimal verbal exchange. When the subject indicated that a reading was finished, the investigator pressed a key, the previous image was erased from the screen, and the next image appeared, scrolling from the top of the screen to the bottom. The average time between images was 5 seconds and it took approximately 9 seconds for the entire image to appear on the screen.

Once all ten cases had been read, and the session was finished, the subjects were asked informally to state any overwhelming impressions they may have had with regard to the computer images. This was mainly to detect any serious problems that might have adversely affected the entire study. No such difficulties were expressed, and the overall impression of the images was very favorable.

## 4 Analysis of Data

The data collected for this study consisted of transcripts of the dictations that the subjects made while viewing the cases, together with tim-

ing information taken from the videotapes of the sessions. A dictation transcript for each case (film-only) was also provided by the radiologist member of the investigation team. These latter dictations were used as a baseline for a subjective evaluation of performance. Thus, three different categories of data were analyzed. The first evolved from the subject transcripts and concerned feature-related data such as landmark references, requests for additional views or studies, comments on image quality, and specific size references. The second category involved an analysis of the time taken to read an image, while the third category resulted in an assessment of some performance trends. All analyses were tabulated on a per-subject basis and on a per-case basis. Film and CRT results were separately tabulated, and a further differentiation was made between overall subject performance and expert/resident performances.

### 4.1 Feature Analysis

Verbal references to features in the images were tabulated from the transcripts. These references were entirely spontaneous during the dictation process, and thus some subject-to-subject variation was encountered. The types of features considered included anatomical landmarks (i.e., common anatomical references that were not necessarily diagnosis-related), spatial references (i.e., vertical - above, below, etc.; horizontal - left, right, etc.; out-of-image-plane - front, behind, etc.), requests for additional information (other views, previous studies, other modalities, etc.), comments on image quality, and size assessments.

Ten basic landmarks were mentioned most frequently by the subjects across all the cases: lungs, heart, thorax/bones, mediastinum, hilum, heart border, costophrenic angles, trachea, vasculature, and aorta. Figure 3 shows the percentage reference to each of these physiological landmarks for all subjects over all test cases. It is interesting to note that there was little variation between the numbers for either film or CRT. Residents tended to articulate a bit more than experts (Figs. 4 and 5). However, this observation may be skewed by the fact that there were only two expert subjects, one of whom consistently gave terse, undetailed dictations. Vasculature was always mentioned less frequently for CRT, which leads us to question the visibility of vascular structures on the CRT display.

The number of horizontal references overall dominated the spatial references, while vertical and out-of-image-plane references occurred with approximately the same frequency (Fig. 6). This may be related to the large amount of horizontal symmetry that occurs anatomically.

Requests for additional information occurred far more frequently for residents than for experts, a fact which may be related to greater confidence of the experts in their diagnosis. An interesting related observation is the fact that CRT readings prompted markedly more requests for additional information in both experts and residents, as compared to film readings (Fig. 6).

Comments on image quality (e.g., overpenetrated, underpenetrated, etc.) were tabulated and while virtually no comments were made with respect to film images, a marked number of comments occurred against the equivalent CRT images (Fig. 6), most frequently referring to underpenetration or brightness of the images. A marked proportion of the comments were made by residents as opposed to experts. However, 8 out of 10 subjects made some comment about the CRT image quality at least once. An interesting observation with respect to this data is that all comments about CRT image quality occurred within the first four images, even though all images were processed in the same way. This leads us to consider whether some of the comments might have been the result of an "accommodation" effect of looking at CRT images, or whether the images were truly of poor quality.

In some cases the subjects mentioned specific sizes of anatomical objects such as nodules or masses. A tabulation of these references according to the difference from the mean for each film shows that the range of CRT size variation is somewhat greater than that of film (Fig. 7). Individual performance comparisons also show a tendency to overestimate size on the CRT image as opposed to the same reference on the film image. However, these results may not be conclusive due to the somewhat small sample size (five subjects and five films).

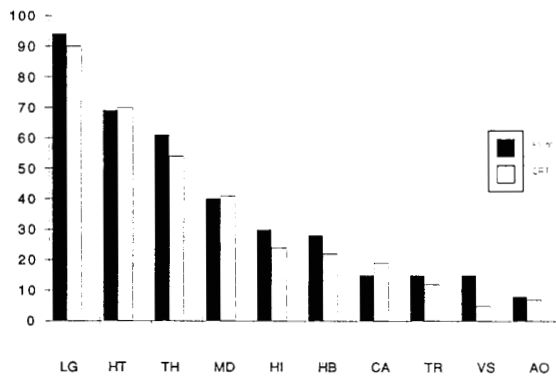


Figure 3: Landmark References over 10 Test Cases, All Subjects (LG - Lungs, HT - Heart, TH - Thorax/Bones, MD - Mediastinum, HI - Hilum, HB - Heart Border, CA - Costophrenic Angles, TR - Trachea, VS - Vasculature, AO - Aorta)

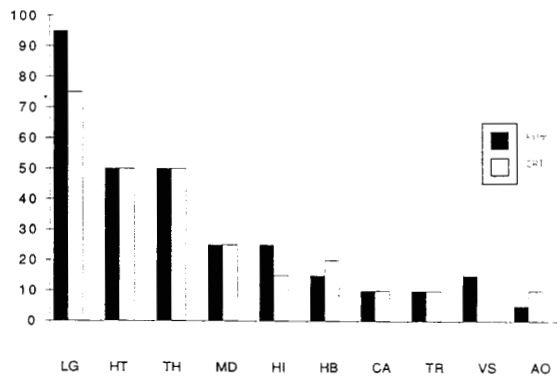


Figure 4: Landmark References, Experts Only

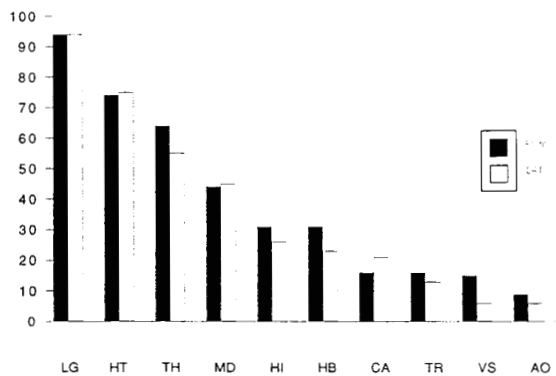


Figure 5: Landmark References, Residents Only

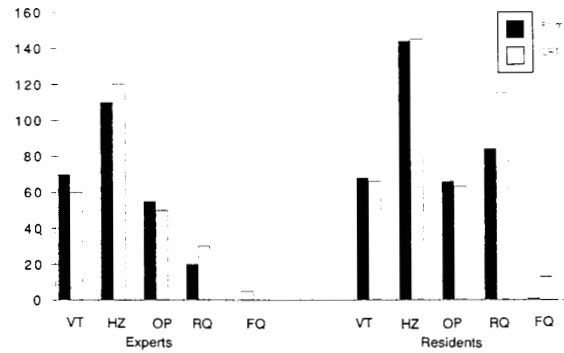


Figure 6: Spatial References, Other Requests, Image Quality (VT - Vertical Refs, HZ - Horizontal Refs, OP - Out-of-Plane Refs, RQ - Other Requests, FQ - Image Quality)

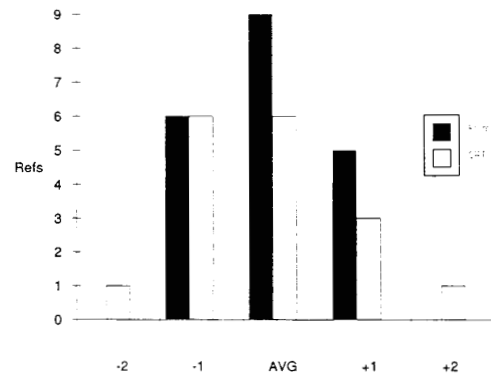


Figure 7: Size Comparisons for Film and CRT Images

## 4.2 Timing Data

Figure 8 depicts the average timings for the studies. The experts performed as expected with a higher throughput than the residents. It is also interesting to note that this discrepancy was media independent (i.e., film or CRT) and that the variation on interpretation time between film readings and CRT readings was insignificant.

It is more difficult to generalize on an individual test case basis regarding the variation in timings between film and CRT readings. In the case of experts, 40 percent of the readings were markedly slower using CRT while 20 percent were faster. For residents, the results were 60 percent markedly slower using CRT with 30 percent faster. When comparing expert versus resident reading times for film only, we found that in 60 percent of the cases, the residents took markedly longer to interpret the data while the other 40 percent had very similar timing performance (Fig. 9). Experts fared even better on CRT cases with 90 percent of the cases interpreted more rapidly than the residents (Fig. 10).

## 4.3 Performance Data

Diagnostic performance was rated on a scale from 0 to 2, where the subjects' primary diagnoses were compared to those of the radiologist investigator as described in a previous section (Figs. 11 and 12). A score of 0 was assigned for complete disagreement between diagnoses, a score of 1 was given if no diagnosis was articulated, or if the diagnosis was more general than the standard. A score of 2 meant complete agreement with the diagnosis. These scores were subjectively assigned with the aim of obtaining some preliminary data regarding performance behavior. Overall, the residents failed to perform as well in diagnostic accuracy when compared to experts in 80 percent of the film cases and in 70 percent of the CRT cases. The experts showed a marked similarity of performance between CRT and film in 60 percent of the cases. In only 1 case was expert performance on CRT better than that of film. On the other hand, residents only achieved similar performance in both media in 1 case, but showed better performance on CRT in 50 percent of the cases. In fact, the residents performed better than the experts on 30 percent of the CRT readings.

## 5 Conclusions and Future Work

With the completion of the Observation Phase data collection, we are well prepared to design the next phase of the research study, namely the Intrusive Phase. The three goals of the transitional part of the Observation Phase have been met. In particular, our subjects are now acclimatized to the environment of the Imaging Sciences Laboratory, and we believe that this change in environment from reading room to laboratory will not adversely affect the performance of the subjects in future experiments. Secondly, the subjects have been exposed to the CRT-displayed X-ray images and, beyond some comments on the brightness of the images, have expressed no concern about their ability to work with these images. Finally, a comparison between the film and CRT protocols shows sufficient similarity in feature references, timings and subjective performance ratings to allow us to proceed in the design of the Intrusive Phase, which will involve CRT X-ray images only.

One of the objectives for the Intrusive Phase is the development of a more formal approach to extracting information from the radiologists, while maintaining minimum interference with their reasoning process. One way of accomplishing this is to allow each subject to dictate the findings for each image, and during the dictation, to pose some very specific questions about what is seen and what is said. These questions are designed to extract more detail about what is seen rather than to force introspection upon the subject. Subjects will also be encouraged to "talk out loud" while they are examining the images, again, mainly for descriptive and elaborative purposes. After all the images for the session have been read, another set of questions will be presented, this time, requesting that the subjects describe any mental representations and/or processes of which they are aware, either while reading images, or in retrospect. At this stage, the subjects will be encouraged to "reason about their reasoning" and to talk about what kinds of strategies

they believe they are using.

Of the three categories of data which were described in the Analysis section, namely, features, timing and performance, only features and performance are relevant in the Intrusive Phase, since the overall film reading times will be affected by the interaction between investigator and subject. The need for a lexicon of both anatomical landmarks and diagnoses understood and accepted by experts and residents alike, has become evident from the work in the Observation Phase. The development of such a lexicon will be undertaken in the Intrusive Phase in order to reconcile the terminology used by different subjects, and to make these terms accessible for the later development of a knowledge-based image enhancement system.

## 6 Summary

The results presented in this paper represent the outcome of the first phase, referred to as the Observational Phase, of a larger, three-stage research study. This Observational Phase included both a Reading Room component, whereby the regular activities and behaviors of radiologists were observed on a daily basis, and a Transitional component, during which selected subjects were observed performing radiological diagnosis on cases presented both on film and on a computer screen. The results of the transitional component were tabulated and analyzed according to features, timing and performance. No significant problems were encountered in the transition from film to CRT, and it has been shown how the results from this part of the study will be utilized to develop the experiment for the second stage, called the Intrusive Phase. The main goal of the study is to develop a cognitive model of the radiological diagnostic process that will be useful for developing a knowledge-based visualization system. The third stage of the study, called the Enhancement Phase, will test the validity of the model in its prediction for pertinent image enhancements.

## Acknowledgements

This research was supported by the Emory/Georgia Tech Biotechnology Research Center. In addition, the authors would like to thank the following people for their assistance to this project: the staff of the Division of Imaging Physics of Georgetown University Medical Center, and in particular, Mr. Robert E. Braudes; Mr. David Cooke, Mr. Marijn Brummer, Mrs. Annie Annis and the staff of Mr. Barry Glasser, all of the Department of Radiology at Emory University Hospital.

## References

- [1] Ahmed, W., Turner, and Hall, E. Quantitative Diagnosis of Early Pulmonary Edema. *Proceedings of the Symposium on Computer-Aided Diagnosis of Medical Images*. Long Beach, CA: IEEE, 1976, pp. 9-13.
- [2] Baker, K.D. and Lloyd, R.O. The Application of Multiple Bandpass Filters to Computed Tomographic (CT) Images. In C. Carl Jaffe (Ed.), *2nd International Conference on Visual Psychophysics and Medical Imaging*. New York: IEEE, 1981, pp. 165-172.
- [3] Blesser, B. and Ozonoff, D. A Model for the Radiologic Process. *Radiology* 103, June 1972, pp. 515-521.
- [4] Christensen, E., Murry, R., Holland, K., Reynolds, J., Landay, M. and Moore, J. The Effect of Search Time on Perception. *Radiology* 138, February 1981, pp. 361-365.
- [5] De Valk, J.P.J. and Eijkman, E.G.J. Feasibility of Features for Visual Classification of Lung Radiographs. In van Bommel, Ball and Wigertz (Eds.), *MEDINFO-83*. Amsterdam: North Holland, 1983, pp. 337-340.
- [6] Fiske, R., Valentino, D., Huang, H., and Blume, H. The Effect of Digital Image Display Format on Perceived Image Quality. *SPIE Vol. 767 Medical Imaging*, 1987, pp. 631-638.
- [7] Hendee, W. and Wells, P. Visual Perception as an Opportunity for Radiologic Research. *Investigative Radiology* 24, July 1989, pp. 575-576.

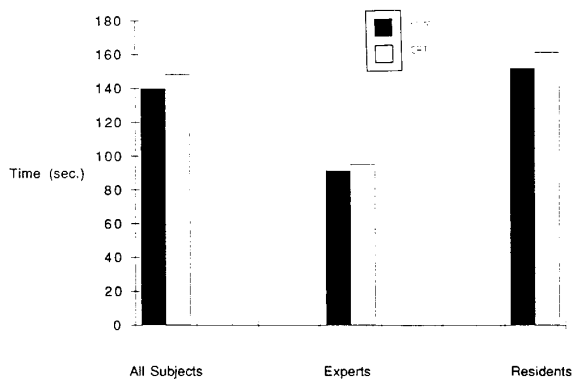


Figure 8: Average Timings Over 10 Test Cases

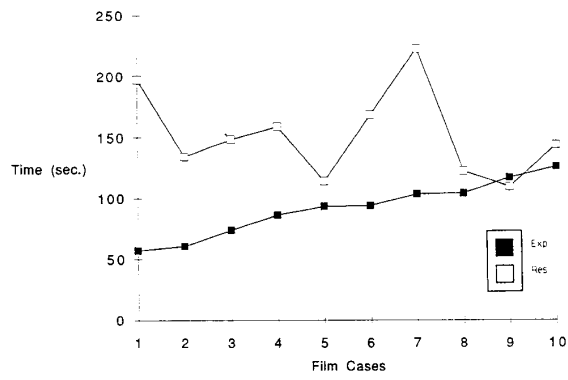


Figure 9: Expert vs. Resident Film Timings

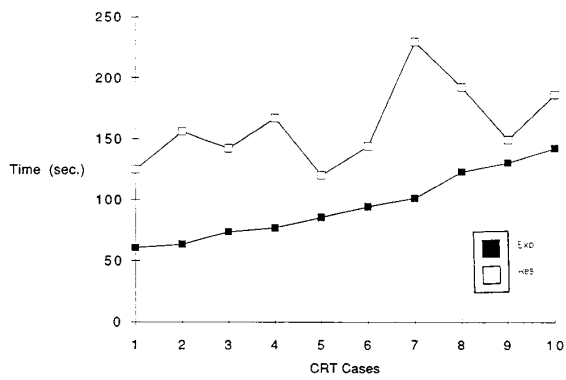


Figure 10: Expert vs. Resident CRT Timings

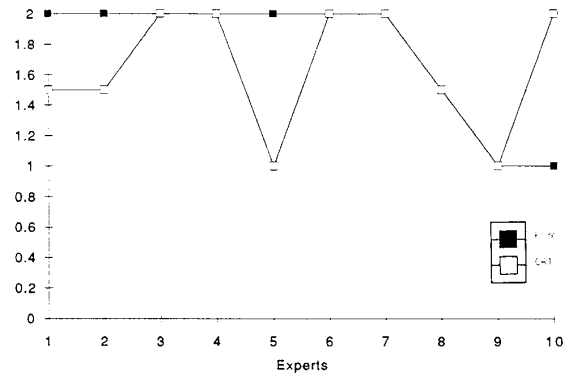


Figure 11: Subjective Performance Evaluation for Experts

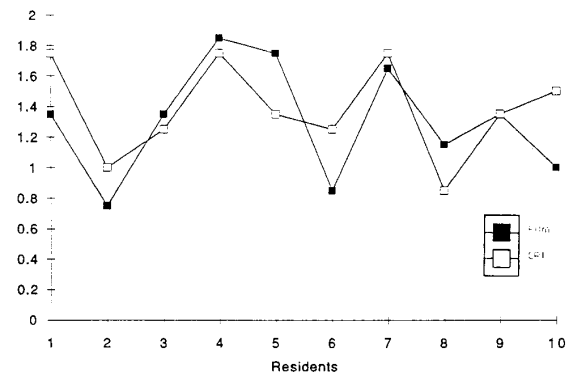


Figure 12: Subjective Performance Evaluation for Residents

- [8] Katsuragawa, S., Doi, K. and MacMahon, H. Quantitative Analysis of Lung Texture for Computer-Aided Diagnosis of Interstitial Disease in Digital Chest Radiographs. *SPIE Vol. 914 Medical Imaging II*, 1988, pp. 371-374.
- [9] Kimme, C. and Sklansky, J. Medically Significant Features for Image Display. *Proceedings of the Symposium on Computer-Aided Diagnosis of Medical Images*. Long Beach, CA: IEEE, 1976, pp. 53-59.
- [10] Kundel, H., and Nodine, C. Interpreting Chest Radiographs Without Visual Search. *Radiology* 116, September 1975, pp. 527-532.
- [11] Lesgold, A., Feltovich, P., Glaser, R. and Wang, Y. The Acquisition of Perceptual Diagnostic Skill in Radiology. *Learning Research and Development Center Technical Report No. PDS-1*, September 1981.
- [12] Ligtenberg, A., Biemond, J. and van der Wildt, G. Optimal Restoration of Noisy Images for the Human Observer. In C. Carl Jaffe (Ed.), *2nd International Conference on Visual Psychophysics and Medical Imaging*. New York: IEEE, 1981, pp. 125-130.
- [13] Marr, D. and Nishihara, H. Visual Information Processing: Artificial Intelligence and the Sensorium of Sight. In M. Fischler and O. Firschein (Eds.) *Readings in Computer Vision*. Los Altos, CA: Kaufmann, 1987.
- [14] Moseley, R.D. and Kelsey, C.A. Observer Performance Testing in Clinical Radiology. In C. Carl Jaffe (Ed.), *2nd International Conference on Visual Psychophysics and Medical Imaging*. New York: IEEE, 1981, pp. 103-108.
- [15] Myles-Worsley, M., Johnston, W. and Simons, M. The Influence of Expertise on X-Ray Image Processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 14, No. 3, 1988, pp. 553-557.
- [16] Piotrowski, L. The Contributions of Spatial-Frequency Amplitude and Phase Information for Visual Processing. In C. Carl Jaffe (Ed.), *2nd International Conference on Visual Psychophysics and Medical Imaging*. New York: IEEE, 1981, pp. 49-51.
- [17] Revesz, G., and Kundel, H. Psychophysical Studies of Detection Errors in Chest Radiology. *Radiology* 129, June 1977, pp. 559-562.
- [18] Seeley, G., Robles-Sotelo, E., Cannon, G., Bjelland, J., Ovitt, T., Standen, J., Capp, M., Fisher, H., and Dallas, W. The Use of Psychophysics as a System Design Aid: Comparison of Film-Screen to an Electronic Review Console. *SPIE Vol. 767 Medical Imaging*, 1987, pp. 639-642.
- [19] Seeley, G., and Newell, J. The Use of Psychophysical Principles in the Design of a Total Digital Radiology Department. *Radiologic Clinics of North America* 29, No. 2, June 1985, pp. 341-348.
- [20] Sezan, M., Tekalp, A. and Schaetzing, R. Automatics, Anatomically Selective, Artifact-Free Enhancement of Digital Chest Radiographs. *SPIE Vol. 914 Medical Imaging II*, 1988, pp. 364-370.
- [21] Sharp, P., Chesser, R. and Dendy, P. Pixel Size and the Quality of Clinical Radionuclide Images. In C. Carl Jaffe (Ed.), *2nd International Conference on Visual Psychophysics and Medical Imaging*. New York: IEEE, 1981, pp. 81-86.
- [22] Sklansky, J., Sankar, P., Katz, M., Towfiz, F., Hassner, D., Cohen, A., and Root, W. Toward Computed Detection of Nodules in Chest Radiographs. In *Proceedings: Sixth Conference on Computer Applications in Radiology and Computer-Aided Analysis of Radiological Images*. Long Beach, CA: IEEE, 1979, pp. 249-252.
- [23] Starr, S., Metz, C., Lusted, L., and Goodenough, D. Visual Detection and Localization of Radiographic Images. *Radiology* 116, September 1975, pp. 533-538.
- [24] 't Hoen, P.J. The Influence of Unsharpness on the Threshold Visibility of Edges, Bars and Disks. In C. Carl Jaffe (Ed.), *2nd International Conference on Visual Psychophysics and Medical Imaging*. New York: IEEE, 1981, pp. 53-57.
- [25] Tuddenham, W. Visual Search, Image Organization, and Reader Error in Roentgen Diagnosis. *Radiology* 78, May 1962, pp. 694-704.
- [26] van der Wildt, G.J. and Waarts, R.G. The Influence of the Surround on the Contrast Sensitivity. In C. Carl Jaffe (Ed.), *2nd International Conference on Visual Psychophysics and Medical Imaging*. New York: IEEE, 1981, pp. 59-63.
- [27] Webb, S. Significance and Complexity in Medical Images - Space Variant Texture Dependent Filtering. In C.N. de Graaf and M.A. Viergever (Eds.) *Information Processing in Medical Imaging*. New York: Plenum Press, 1988.
- [28] Yasuda, Y., Kinoshita, Y., Emori, Y., and Yoshimura, H. Computer-Aided Diagnostic System for Mass Survey Chest Images. *SPIE Vol. 914 Medical Imaging II*, 1988, pp. 772-779.